[1] **Gulmira Bekmanova**

[2] **Banu Yergesh**

[3] **Assel Omarbekova**

[4] **Altanbek Zulkhazhav**

The Intelligent System for Low Recourses Languages for Expanding Language Recourses

Abstract: - This research, funded under the BR11765535 “Development of Scientific and Linguistic Foundations and IT Resources to Expand the Functions and Improve the Culture of the Kazakh Language” project, is aimed at the need to develop the scientific and linguistic foundations of the Kazakh language with a special focus on anthroponymy. The "Akhmettanu" intelligent system, which includes digitized materials based on the scientific heritage of Akhmet Baitursynuly, consists of an ontological model of knowledge on the structural layers of the language, a Kazakh language user interface with the function of adding and correcting new digitized materials. Digitized materials based on the scientist's work "Language-Tool" include a new system of knowledge about the sound, word and sentence system of the language. The main aim of this system is to promote the scientific heritage of Akhmet Baitursynuly by promoting the further development of digitalized knowledge and information tools that affect the development of the Kazakh language.

Keywords: Intelligent system, Kazakh language, Ontology, Semantic knowledge base.

I. INTRODUCTION

This work was funded in the framework of the research project BR11765535 “Development of Scientific and Linguistic Foundations and IT Resources to Expand the Functions and Improve the Culture of the Kazakh Language”. The goal of the project is to develop scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language as a language of interethnic communication in a digital format, which is an urgent and important problem in the strategic direction of the development of our state. The main idea of this project is to solve the tasks set using the methods of linguistics and artificial intelligence through a formalized and semantic description of the Kazakh language, the construction of predicative knowledge bases and software implementation of the obtained models and methods.

As a result of the research, grammatical rules of the Kazakh language were generalized, systematized and described in the predicate language, a model of the Kazakh speech synthesis system was developed, phonological analysis was carried out and a transcriber of the speech synthesis system based on A. Baitursynov’s teachings was developed, an ontological model based on Akhmet Baitursynov’s legacy for the intellectual system “Akhmettanu”, an acoustic corpus based on A. Baitursynov’s work “Til – Kural” was developed.

As a prerequisite for the creation of the intellectual system "Akhmettanu", A.Baitursynuly's entire legacy was collected and classified into several categories in terms of content and subject matter. Textbooks, alphabets, and articles were divided into 5 sections according to the topic: language, literature, methodology, other scientific fields, and people's heritage. And those departments were further systematized within themselves. For example, the "Language science" section included textbooks, alphabets, and articles related to language education of the teacher of the nation, and the "Literary science" section included textbooks, poems, translations, and articles related to the field of literature. The "Methodological Science" department has grouped teaching-methodological tools and articles related to teaching, learning and education issues. The "People's Heritage" section includes songs collected by the scientist A. Zataevich and poems such as "Er Sayin", "23 lamentations". The bibliographic list of all works of the scientist included in these sections was systematized and electronic versions were collected.

Within the framework of research work A. Baitursynuly's work "Language-tool" [1,2] presented the terms related to the sound system and their definitions, sound types, the division of vowels and consonants, their characteristics, syllables, and the basis of the transport system.

Operation was done on the "Word System" section of A. Baitursynuly's "Language-Tool" work. Based on the

[1,2,3,4] L. N. Gumilyov Eurasian National University

[1]gulmira-r@yandex.kz, [2]b.yergesh@gmail.com, [3]omarbekova@mail.ru, [4]altinbekpin@gmail.com

morphological system of the Kazakh language, drawings were prepared as a basis for creating models. That is, the theoretical data related to the classes of words and persons of speech included in the group of nouns and epithets were systematized and drawn. The drawings are organized in several parts, such as the name of the concepts, their definition, for example, their division, transformation, tasks for training.

A database of terms and their definitions, parts of sentences, types of sentences related to the sentence system proposed by A.Baitursynuly was created.

Language is a systematic phenomenon. The word is the most important unit of the language system. The importance of a word as a linguistic unit is determined, first of all, by its participation in language levels. In this regard, there is also a basis for differentiating the conclusions of the founder of Kazakh linguistics A. Baitursynuly about the word and its system, the personality of the word, classes of words.

Research on the grammatical structure of the Kazakh language, including the personality of the word, classes of words, the system of word transformations, originates from the scientific reasoning of A. Baitursynuly.

A. Baitursynuly accurately understood the nature and personality of the language, including the Kazakh language, and studied the structural levels of the language in accordance with it. In determining the features, patterns and concepts of the language in relation to the sound, word, sentence and speech system, typological features of the structure of the Kazakh language were taken as a basis.

II. ONTOLOGICAL MODELS OF ALL STRUCTURAL LEVELS OF THE LANGUAGE ACCORDING TO THE WORKS OF AKHMET BAITURSYNULY

There is a need to model the legacy of Akhmet Baitursynuly, the scientist's works on the basis of a systematized knowledge base for the purpose of digitizing and further digital processing of knowledge related to all structural layers of the language. In this work, an ontological modeling tool was taken to model the knowledge in question.

For the intellectual system "Akhmettanu" a knowledge base was created based on the heritage of Akhmet Baitursynuly, and an ontological model of the structural levels of the language was created.

Modeling is a powerful and effective tool for research and development of intelligent systems for various subject areas.

One of the ways of creating intelligent systems is ontological modeling.

Ontology is a conceptual drawing consisting of a set of concepts and a set of statements about these concepts. Based on it, classes, relations, properties, functions and individuals (instances) can be described [3]. This is a form of formal representation of a certain field of knowledge. Currently, ontology is widely used in programming, teaching, and various researches.

According to the works of Akhmet Baitursynuly, the ontological model of the structural levels of the language was created in the environment of Protégé [4] This is because this environment allows us to describe concepts and concrete objects. It also has many sets of operators. They consist of descriptions of classes and their instance versions (individuals). By creating such an ontology, it is possible to obtain logical consequences.

According to the works of Akhmet Baitursynuly, the ontological model of the structural levels of the language consists of classes (concepts), their properties and individual models (individuals).

Ontological models of the structural levels of the language, including the sound system, person of speech, word classes, sentence types, were created based on the works of Akhmet Baitursynuly.

The ontological model of the sound system proposed by A. Baitursynuly is shown in Figure 1.

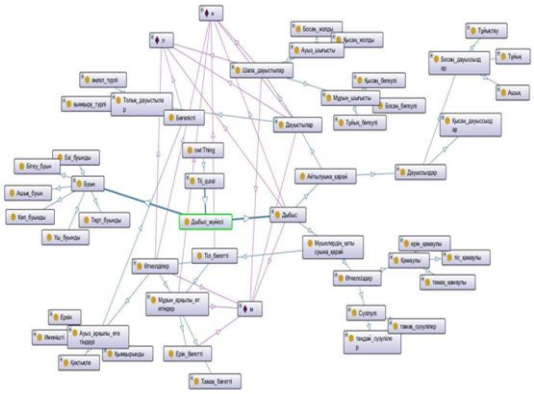

Figure 1 - An excerpt from the ontological model of the sound system proposed by A. Baitursynuly

The ontological model of the system proposed by A. Baitursynuly describes the classification of word forms and word classes and their transformation (**Error! Reference source not found.**).

III. THE INTELLIGENT SYSTEM "AHMETTANU"

General information

The intelligent system "Akhmettanu" allows to systematize the scientific heritage of Akhmet Baitursynuly and to get an answer to the question based on the ontological model formed by the new knowledge system with all the structural layers of the language according to his studies.

The "Akhmettanu" intelligent system, which includes digitized materials based on the scientific heritage of Akhmet Baitursynuly, consists of an ontological model of knowledge on the structural layers of the language, a Kazakh language user interface with the function of adding and correcting new digitized materials. Digitized materials based on the scientist's work "Language-Tool" include a new system of knowledge about the sound, word and sentence system of the language.

The main goal of this system is to promote the scientific heritage of Akhmet Baitursynuly by promoting the further development of digitalized knowledge and information tools that affect the development of the Kazakh language.

The user of the intellectual system "Akhmettanu" gets acquainted with the scientific works of A. Baitursynuly and receives the necessary information about the main concepts and concepts presented by the scientist according to the levels of the Kazakh language, as well as scientific judgments.

Technical requirements and description of the operating system for the server device on which the intelligent system "Ahmettanu" will be installed:

Type of computer running: 2.8 GHz 4-core processor, 4 GB of RAM, 50 GB of hard disk,

Operating system: Linux (Debian), Windows 7 and higher

Software: Web browser of the latest version: Google Chrome, Firefox, Opera, etc. p.; Added support for javascript and cookies.

Programming language: Python.

Creation of the Frontend section of the intelligent system

In the question-and-answer section of the intelligent system, it is possible to get answers to questions based on the ontological model of the structural levels of the Kazakh language based on the work of the scientist "Language-tool".

The Frontend part of the project is implemented in Vue.js.

Technologies and tools used:

Vue.js. was used to create the client part of the application, including the administrator interface for processing information in the system. Vue.js allows you to create dynamic and fast-loading web pages.

Axios. A library for making HTTP requests.

Vuetify: A library for improving UI.

Vuex: State management library.

Creation of the Backend section of the intelligent system

The backend part of the question and answer section of the intelligent system was implemented in Flask.

Technologies and tools used:

Flask. The Backend part of the project was created in this Python framework. Flask was chosen for its simplicity and flexibility, which allowed us to quickly deploy an API to interact with the client side. It is also extensible thanks to the support of various plugins and libraries.

Main functions:

Handling HTTP requests from the frontend.

Interaction with databases and analysis algorithms.

Used semantic technology: Owlready. Owlready library is used to work with semantic networks and ontologies. It allows OWL ontology integration with Python code and Sparql queries, making morphological and semantic analysis more accurate.

System architecture and deployment.

The system is deployed as microservices, each of which runs in a separate Docker container:

PostgreSQL: A standalone Docker container for databases. Contains all necessary tables and relationships.

API: Implemented in Flask and run in its own Docker container.

UI: The client side is written in Vue.js and runs in its own Docker container.

NGINX: Used as a reverse proxy to route requests between client and server services.

All of these services are enabled using Docker Compose, which provides convenience and automation of the

deployment process.
In order to get an answer to a question in the intelligent system, it is possible to return an answer to a question based on an ontological model, to automatically fill the database of questions and answers, to search for questions by meaning.
Operation of the intellectual system "Akhmettanu"
To use the system, go to https://kazlangres.enu.kz/ on a computer or phone browser. The "Ahmettanu" intelligent system opens in the main window. On the left side of the window, in the "Akhmettanu" menu, you can go to the sections such as "Personality of the scientist", "Heritage", "Language-tool", "Question-answer system", "Terms and definitions" (Figure).

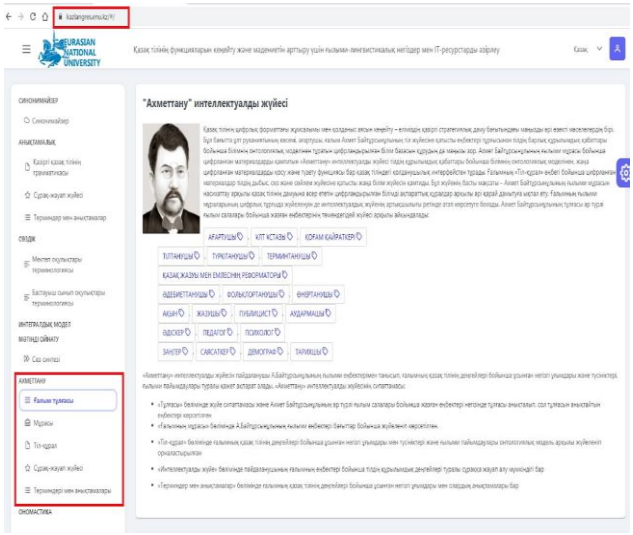

Figure 2 – Main page of the intelligent system "Akhmettanu "

User group:
System Administrator;
Expert Specialist
User (guest).
The system administrator adds expert specialists, gives roles, loads files, enters information, changes, adds, changes the section, subsections.
An expert enters information, adds, changes, downloads files, has the ability to correct information, accepts changes submitted by registered users, publishes them, tags works in the scientist's legacy by personality, asks questions and gets answers in the Intelligent System section.
The user (guest) can view the information in the system, get acquainted with the heritage of the scientist, ask a question and get an answer in the question and answer system section.
The intellectual system " Akhmettanu " consists of the following parts:
In the "Personality of a Scientist" section, the personality of Akhmet Baitursynuly is determined based on the system description and the works written by Akhmet Baitursynuly in various fields of science, and the works defining that personality are shown. Link: https://kazlangres.enu.kz/#/ .
In the "Heritage" section, the scientific works of A. Baitursynuly are systematized by direction. Link: https://kazlangres.enu.kz/#/legacy .
In the "Language-tool" section, the main concepts and understandings presented by the scientist on the levels of the Kazakh language are systematized using the ontological model. Link: https://kazlangres.enu.kz/#/intelect/2 .
In the "Question-Answer System" section, the user has the opportunity to get answers to questions about the structural levels of the language based on the scientist's works. Link: https://kazlangres.enu.kz/#/answer/question/2 .
The "Terms and definitions" section contains the main concepts and their definitions proposed by the scientist according to the levels of the Kazakh language. https://kazlangres.enu.kz/#/terminology/1.

CONCLUSION

In conclusion, during the creation of the base, work was carried out on the analysis, differentiation of the grammatical structure of the Kazakh language, systematization by structural levels of the language based on the works of A. Baitursynuly. In this way, formed the terms of the sound system, word system and sentence system of

the language and introduced them into scientific circulation. Expanded access to A. Baitursynuly's legacy and created an opportunity to create a unified system from heterogeneous blocks of information.
The following results were obtained during the research work.
The basis of terms and their definitions, types of sounds, division of vowels and consonants, character, syllable and transport system proposed by A. Baitursynuly was created:
A system of terms related to the word system and their definitions, person of speech, classes of words, their division, transformation was compiled;
Drawings were prepared as a basis for creating models based on the morphological system of the Kazakh language written in the "Word system" section. The theoretical data related to the classes of words and persons of speech included in the group of nouns and adverbs were systematized and drawn. The drawings were systematized in several parts, such as the name of the concepts, their definition, for example, their division, transformation, tasks for training;
A digital source aimed at systematizing the scientific heritage of A. Baitursynov was supplemented with the following information:
a) language science: textbooks, alphabets;
b) Literary Studies: textbook, poems and translations;
c) methodology: teaching aids, articles;
d) Folk Heritage: original works and songs, songs collected by the poet from the people;
e) by sound system: vowel, consonant, syllable, transport, their definition, classification, types, examples;
f) writing principles, rules;
g) suffixes according to the word system, their types, usage.
Digitized materials on the structural levels of the language based on the work of Akhmet Baitursynuly “Language-tool”;
Knowledge base on the structural levels of the language on the legacy of Akhmet Baitursynuly;
The structural levels of the language proposed by Akhmet Baitursynuly were modeled ontologically;
Based on the works of Akhmet Baitursynuly, based on the knowledge base, ontological models of the structural levels of the language, an intelligent system "Akhmettanu" has been developed, which includes digitized materials based on the scientific heritage of Akhmet Baitursynuly and a new knowledge system with structural layers of the language based on his studies.
Assessment of the completeness of solutions to the obligations: the entire planned report in accordance with the calendar plan has been completed in full.
The results of research work were tested at international conferences and published in scientific publications. Also, according to the works of Akhmet Baitursynuly, an author's certificate was obtained for an applied ontological model of structural levels of the language (certificate № 36969, 09.06.2023 authors: Ergesh B. zh., Bekmanova G. T., Syzdykova G. O., Zhumagulova A. A.). One of the chapters of the collective monograph "development of scientific linguistic foundations and IT resources to improve the culture of the Kazakh language and expand functions", published as part of the implementation of the program, is devoted to the development of the intellectual system "Akhmettanu".
Recommendations and initial data on the actual use of research results: the knowledge base of the structural levels of the language based on the works of Akhmet Baitursynuly, the developed ontological model was used in the development of The Intelligent System "Akhmettanu", which includes digitized materials formed from the new knowledge system on the scientific heritage of Akhmet Baitursynuly. The intelligent system can be used by the general public.
Due to the fact that the implementation of the results obtained during this period was not considered, an assessment of the technical and economic efficiency of the implementation was not carried out.
Until now, the knowledge base of the structural levels of the language based on the works of Akhmet Baitursynuly has not been developed, systematized, and modeled in open use in this field, and the system for processing the obtained results has not been developed. It was developed for the first time within this program.

REFERENCES

[1] Baitursynuly A. Til – kural. - Almaty: Sardar, 2009. - 348 bet. (In Kaz.)
[2] Baitursynul A. Til tagylymy. - Almaty: Ana tili, 1992. - 448 b. (In Kaz.)
[3] Gruber T. R. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2):199-220. – 1993. (In Eng.)
[4] Protégé. – URL: https://protege.stanford.edu/. (karalgan kuni: 10.008.2023). (In Eng.)

AUTHORS PROFILE

Gulmira Bekmanova received the B.S. degree in information systems from D. Serikbayev East Kazakhstan State Technical University, Ust-Kamenogorsk, Kazakhstan in 2004 and the M.S. degree in information systems from the same University in 2006, candidate of technical sciences on specialty: 05.13.11 – Mathematical software for computers, complexes and computer network and the Ph.D. degree in Computer science, computer engineering and control from L.N. Gumilyov Eurasian National University, Kazakhstan in 2010.

From 2004 to 2006, she was Assistant of Professor with the D. Serikbayev East Kazakhstan State Technical University, Kazakhstan. Since 2006 to 2019 she was senior lecturer, associate professor, Researcher of Artificial Intelligence Research Institute, head of the Computer Science and Information Security Department, L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2019 she was the director of the Distance learning center, L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2020 she is the director of the Digital Development and Distance learning Department, L.N. Gumilyov Eurasian National University, Kazakhstan She is the author of 5 monographs, 4 manuals and more than 100 articles, and 2 inventions.

Assel Omarbekova graduated with honors from the S. Toraighyrov Pavlodar State University. Kazakhstan in 2000 and postgraduate from L.N. Gumilyov Eurasian National University, Kazakhstan by specialty Mathematical and software of computers, complexes and computer networks in 2000. In 2008 she was received the degree of Candidate of Science.

Since 2020 she is the deputy director of the Digital Development and Distance learning Department, L.N. Gumilyov Eurasian National University, Kazakhstan

Her research interests related with intelligent tutoring systems, and distance learning technologies. She has more than 250 scientific and scientific-methodical works, of which 24 publications are indexed by Scopus (of which 3 are in peer-reviewed scientific publications with a CiteScore percentile in the Scopus database of at least 35) and 14 - Web of Science, 28 - in publications recommended by CQASES, 10 textbooks, 23 study guides.

She is the author of the following software: technology of automatic create electronic textbooks, system electronic questionnaire of teachers and etc. Also she was involved in projects related to the creation of multimedia learning systems for the Ministry of Education and Science, Ministry of Culture and Information, the Ministry of Finance of Kazakhstan and etc.

Altanbek Zulkhazhav pursued his master's degree from 2014 to 2016 at L.N. Gumilyov Eurasian National University in Astana, Kazakhstan. Currently a Ph.D. candidate.

Since 2014, he has been actively engaged in research within the realms of computer linguistics, nlp, knowledge bases, data processing, and the semantic web. H-index (Scopus):2.

He was involved in the following research projects: “Development of electronic thesauri of Turkic languages to create systems for multilingual search and extraction of knowledge" (2018-2019), "Development of scientific and linguistic foundations and IT resources to expand the functions and improve the culture of the Kazakh language" (2022-2023), AP19679847 "Development of methods for the analysis of the Kazakh political discourse" (from 2023)